Author:
Watson, Henry B

Title:
Unrecognised endemic biodiversity within East African Chiloglanis catfish populations

Unrecognised endemic biodiversity within East African *Chiloglanis* catfish populations.

Henry Watson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Masters by Research in the Faculty of Life Sciences.

December 2019

Student Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the *University's Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:

Henry Watson

Date:

08 / 06 / 2020

Supervisor:

Professor Martin Genner, University of Bristol

Dr Christos Ioannou, University of Bristol

Thesis Abstract

Freshwater fish are among the most threatened vertebrates as a result of increased anthropogenic activity in recent years. Despite the extent to which freshwater environments are known to harbour high levels of biodiversity, it is believed that there are large amounts of unrecognised diversity within freshwater fish populations. Strong spatial structure between distinct populations suggests that geographic barriers preventing the dispersal of species and causing reproductive isolation are the drivers behind speciation, specifically within riverine drainage basins. Secondary contact following geological events has also created the need for habitat specialisation and niche occupation within species living in sympatry leading to spatial structure not just between habits but also within them.

Chapter 1 reviews the current literature regarding speciation and drivers of diversity in freshwater environments and examines previous phylogenetic analyses uncovering unrecognised diversity in African catfish. Chapter 2 presents a phylogenetic study into East African catfish belonging to the genus *Chiloglanis* using next generation sequencing techniques to investigate levels of cryptic biodiversity. The results confirm the strong endemic spatial distribution of *Chiloglanis* species consistent with their typical restriction to single catchments, in addition to identifying multiple distinct clades in a region of which there is limited previous knowledge. These results confirm the hypothesis that there is unrecognised endemic diversity within *Chiloglanis* populations and the strict limitations to species ranges suggest that catchment boundaries may act as a geographic barrier preventing gene flow between neighbouring populations and driving the process of speciation.

Contents

List of Figures and Tables

List of Appendices

Chapter One

1.1 – Cryptic Species and Speciation

Speciation and reproductive isolation

The process by which new species arise has been widely discussed since the publication of Darwin's "On the Origin of Species" in 1859 (Palumbi, 1994). The term "speciation" was introduced by Orator F. Cook (1906) and subsequently the mechanisms driving the process have been subject to countless studies across all manner of taxa. The formation of distinct species from a previously homogeneous population is dependent on the existence of some sort of barrier, whether that is a physical or intrinsic, that leads to a restriction in gene flow between members of the same population (Barton & Bengtsson, 1986; Noor et al., 2001; Coyne & Orr, 2004; Lowry et al., 2008; Lin et al., 2009). This reduction of gene flow is known as reproductive isolation (Dobzhansky, 1937; Mayr, 1942; Lowry et al., 2008; Kozak et al., 2012). The importance of reproductive isolation was recognised as a result of work by Theodosius Dobzhansky (1937) and Ernst Mayr (1942), and has been developed further over the last 80 years. Mayr (1963) suggested the separation of reproductive isolation into two broad modes: prezygotic isolation, in which the isolation prevents the fertilisation of eggs, and postzygotic, in which the egg is fertilised but the isolation prevents the formation of fertile offspring (Palumbi, 1994; Coyne & Orr, 2004; Nosil et al., 2005; Moyle, 2007; Kozak et al., 2012). The description and advancement of reproductive isolation has been vital in helping to explain how barriers to gene flow arise, and thereby the mechanisms driving speciation (Palumbi, 1994; Coyne & Orr, 2004; Nosil et al., 2005; Lin et al., 2009).

Cryptic species

Cryptic species are two or more reproductively isolated species that are essentially indistinguishable using morphological characters (Hyde et al., 2008; von der Heyden et al., 2011; Rosser et al., 2019). They provide a unique problem to taxonomists as the identification of taxonomically distinct yet related species is typically achieved via morphological analysis (Bickford et al., 2007). As suggested by Jousson et al. (2000), a large amount of time may pass between initial speciation on a genetic level and the appearance of subsequent morphological variation. As a result of this, it is somewhat probable that the number of cryptic species in existence greatly outweighs any number which may currently be believed (Bickford et al.,

2007; von der Heyden et al., 2011). In fact, Janzen et al. (2017) estimate that somewhere in the range of 10 to 20% of species identified morphologically are, in fact, two or more genetically distinct species. Cryptic species also pose a problem to conservationists, as failures to identify genetically distinct taxa can give rise to issues within species management where they require different strategies to effectively manage populations (Engelbrecht & Mulder, 2000; Hyde et al., 2008). Bickford et al. (2007) suggest incorrect identification of cryptic species may have particularly high importance in the management of large-scale multispecies fisheries, and in the control of agricultural pests.

The evolution of cryptic species can be explained by the mutual benefits of common morphology between species. Shared morphological characteristics, such as colour patterns, may arise from shared evolutionary history, or convergence on an advantageous trait (Armbruster et al., 2015). Predation is a key selective agent acting on morphological traits, and as a result, may influence the formation of cryptic species complexes. Mimicry describes the process by which two or more species converge on shared phenotypic traits as a result of the mutual benefits gained from doing so (Kikuchi and Pfennig, 2013). The presence of sharp dorsal spines and spinal venom in some species of catfish is suggested to have driven the evolution of interspecies mimicry. Both Batesian mimicry, where less harmful or non-harmful individuals mimic the morphology of harmful individuals, and Müllerian mimicry, where harmful individuals evolve to have similar morphology, have been suggested as the mechanisms leading to the cryptic appearance of species of *Corydoras*, a genus of freshwater catfish found across much of tropical Central and South America (Alexandrou et al., 2011; Lima and Sazima, 2017). Anti-predator benefits gained by sharing an appearance with species that are avoided by predators may explain the cryptic appearance of species in catfish more generally.

Examples of cryptic species

Cryptic species have been uncovered across many taxonomic groups. Le Gac et al. (2007) identified the presence of multiple cryptic species within *Microbotryum violaceum*, a species of parasitic fungus that causes the sexually transmitted anther smut disease in the carnation plant family Caryophyllaceae. Despite overlapping geographic ranges across Europe, North America, and South America, the study identified 11 independent lineages many of which were specific to host plant species. Another study by Hebert et al. (2004) investigated the two-barred flasher, *Astraptes fulgerator*, a species of skipper butterfly found throughout North and South

America. Through a combination of morphological, behavioural and genetic analysis, it was determined that what was thought to be a single species is actually at least 10 genetically distinct cryptic species, distinguishable largely by differences in caterpillar morphology and diet.

Cryptic species have also been discovered within megafaunal species. Species complexes have been identified within the giraffe, *Giraffa camelopardalis* (Brown et al., 2007; Petzold and Hassanin, 2020; Winter et al., 2018), and the African elephant, *Loxodonta africana* (Palkopoulou et al., 2018; Roca et al., 2001; Rohland et al., 2010). Studies in the last few years have been able to identify between two and six genetically separate species of giraffe using genetic and phylogeographic analyses (Petzold and Hassanin, 2020; Winter et al., 2018). Initial evidence from Roca et al. (2001) combined morphological, genetic, and geographic variance to distinguish between the African bush elephant, *Loxodonta africana*, and the African forest elephant, *Loxodonta cyclotis* and this distinction has been corroborated by numerous genetic studies of both extinct and extant elephant species (Palkopoulou et al., 2018; Rohland et al., 2010).

1.2 – Freshwater Phylogeography

Phylogeography

Coined by John Avise in 1987, the term 'phylogeography' relates to the historical structure of geographical populations of individuals based on population genetic variance over millions of years (Avise, 1998; Avise, 2000; Searle, 2000; Hewitt, 2001; Knowles & Maddison, 2002; Lemey et al., 2009). Over the last 30 years phylogeography has exploded as an area of evolutionary biology, in part due to its interdisciplinary nature covering subdisciplines including molecular and population genetics, ethology, and historical biogeography (Avise, 1998; Taberlet et al., 1998). As well as its use regarding historical biogeography, Carreras-Carbonell et al. (2005) indicate that phylogeography can be a useful tool in uncovering unknown instances of potential cryptic species. For example, in a study of the red-black triplefin blenny (*Tripterygion tripteronotus*) two distinct populations were identified that were estimated to have diverged 2.75 and 3.32 million years ago during a period of global cooling (Carreras-Carbonell et al., 2005).

Phylogeographic analysis has traditionally largely revolved around the comparison of mitochondrial DNA (mtDNA) among target individuals (Avise, 1998; Hewitt, 2001). This is due to a number of factors. The fast rate of mtDNA evolution, which can be up to 20 times greater than that of nuclear genomic DNA, makes for a highly variable population which can give valuable insight into short-term population structure (Wallace, 1994; Torroni et al., 1996; Finnilä et al., 2000; Galtier et al., 2009; DeSalle et al., 2017). Additionally, highly conserved regions typically surround the regions of variability within mtDNA allowing for primers that are easy to design and effective to use (Galtier et al., 2009; DeSalle et al., 2017). Finally, mtDNA is maternally inherited and does not undergo recombination. This makes phylogenetic analysis much simpler and grants the ability to trace the geographic history of maternal lineages (Galtier et al., 2009; DeSalle et al., 2017). However, comparative analysis of mtDNA and nuclear DNA studies have highlighted some of the drawbacks of using mtDNA in phylogenetic studies (Balloux, 2009; Galtier et al., 2009; DeSalle et al., 2017). The single-locus nature and non-recombining properties of mtDNA can lead to the production of misleading phylogenies due to incomplete lineage sorting (Moore, 1995; Galtier et al., 2009; Choleva et al., 2014). Additionally, the uniparental inheritance of mtDNA can give an incomplete picture of past hybridisation events as hybrids retaining maternally inherited mtDNA can be indistinguishable from purebred individuals following genetic analysis (Freyhof et al., 2005; Galtier et al., 2009; Choleva et al., 2014). Alternatively, rare hybridisation events which may disappear from nuclear genomes within a few generations may be retained in the mitochondrial genome for many generations through maternal inheritance (Freyhof et al., 2005).

The introduction of relatively low-cost nuclear DNA-based next generation sequencing techniques, such as double digest restriction enzyme associated (ddRAD) sequencing, can help to overcome some of the limitations of the use of mtDNA (Sutra et al., 2019). Sequencing ddRAD loci provides information on the states of thousands of loci, thereby limiting the influence of incomplete lineage sorting on any phylogenys reconstructed using the data. Additionally, due to its widespread coverage across the nuclear genome, ddRAD can limit the influence that rare hybridisation events in the past may have (Sutra et al., 2019). Generally, as capacity for genotyping variants within nuclear genomic DNA has increased over recent years using data from high-throughput sequencing methods, there is an increasing reliance on genome-scale evidence for phylogeographic inference.

Phylogeography in a freshwater environment

Phylogeographic studies have been undertaken on a variety of taxa, from viruses to humans, across almost all areas of the globe (Wallace, 1994; Torroni et al., 1996; Avise, 1998; Finnilä et al., 2000; Lemey et al., 2009). The study of freshwater species that cannot typically cross marine or terrestrial environments can provide specific insight how maximum potential ranges are limited by both catchment boundaries, but also intrinsic habitat variability that can inhibit dispersal (Machordom & Doadrio, 2001). Additionally, Bermingham and Martin (1996) point out that phylogeographical analysis of freshwater species provides a tremendous insight into both the biotic and geological history of a location. This is because the distribution of these species is entirely dependent on the historical evolution of the location's hydrographic topography as they can only travel between connected drainage basins, thereby indicating that the phylogeographic study of freshwater populations can provide an insight into the development of a geological region (Bermingham and Martin, 1996; Sivasundar et al., 2001; Dias et al., 2013). The effect of geographical events on freshwater fish speciation, and therefore phylogeography, was studied by Verheyen et al. (1996). Lake Tanganyika underwent a several thousand-year period of aridity roughly 200,000 years ago that caused the water level to drop by 600 metres fragmenting the main water body into three smaller lakes (Tiercelin & Mondeguer, 1991). This lake fragmentation appears to have promoted divergent genomic evolution within shallow water rocky shore species, and that divergence has persisted as the lake levels rose to those of the present data. This is exactly the kind of geographic factor that can help to explain present day phylogeographic structure within a species (Verheyen et al., 1996).

1.3 – Speciation in Rivers

Freshwater systems are an excellent place to examine the mechanisms that lead to a reduction in gene flow between individuals of a population and, as a result, drive the process of speciation. Despite making up only ~0.01% of the Earth's aquatic habitats, freshwater systems are home to over 40% of all known fish species (Lundberg et al., 2000; Lévêque et al., 2008; Bloom et al., 2013; Guo et al., 2019) This clearly indicates the extent of speciation that has occurred and continues to occur within these systems (Lévêque et al., 2008). There are multiple factors that are believed to drive speciation within freshwater, and even more specifically riverine, habitats. When examining the mechanisms behind speciation in this environment it is

important to differentiate between divergence occurring between individual river systems causing allopatric speciation, and divergence occurring within river systems causing sympatric speciation.

A key factor in the divergence found between neighbouring systems is the process of habitat fragmentation (Fuller et al., 2015). Fragmentation in freshwater systems has historically largely been the result of the formation of natural barriers, including waterfalls and beaver dams (Rahel, 2007; Dias et al., 2013; Fuller et al., 2015). Dias et al. (2013) examined the effect of waterfalls as natural barriers to gene flow between freshwater fish species in the Orinoco River basin stretching across Venezuela and Colombia. Analyses of 26 subdrainages indicated that higher levels of fragmentation were directly linked with higher species richness within individual subdrainages. This supported the concept that geographic isolation, caused by natural barriers, acts as a mechanism for speciation within riverine environments (Dias et al., 2013). In recent years however, studies into fragmentation in riverine habitats has focused on man-made barriers such as the building of dams and roads and the introduction of pollutants (Theodorakis et al., 2006; Fuller et al., 2015). A study into the effects of contaminants released by paper mills into Pigeon River, North Carolina on population genetics of the redbreast sunfish (*Lepomis auratus*) showed higher levels of genetic diversity and increased mutation rates when compared with reference populations from uncontaminated areas (Theodorakis et al., 2006). It was proposed that contrasting levels of genetic diversity within sunfish populations are also indicative of patterns of gene flow influenced by man-made contaminants within the river.

Sympatric speciation remains a controversial topic among evolutionary ecologists largely due to the limited number of examples and the restricted nature of the concept (Jiggins, 2006; Foote et al., 2018; Payne and Polechova, 2019). Sympatric divergence, however, has been observed within freshwater systems (Kadye & Moyo, 2008; Seehausen & Wagner, 2014; Malinsky et al., 2015). The mechanisms driving speciation in the absence of any clear geographic boundaries to gene flow often revolve around aspects of behavioural isolation among individuals in the systems. There are numerous biotic and abiotic factors that can promote behavioural isolation within sympatric populations including turbidity, predation regimes, light levels, oxygen levels and trophic preferences (Kadye & Moyo, 2008; Seehausen & Wagner, 2014). The explosive adaptive radiation of crater lake cichlids in East Africa potentially provide an example of real-world sympatric divergence (Malinsky et al., 2015). A study by Malinsky et al. (2015) identified two separate ecomorphs of *Astatotilapia calliptera* from Lake

Massoko in the early stages of divergence whilst remaining in sympatry. Evidence of this divergence mirrors characteristics typically observed in previous examples of cichlid adaptive radiation such as mate preference, morphology, and microhabitat preference (Wagner et al., 2012; Malinsky et al., 2015). However, examples such as this remain controversial as numerous questions can be asked about the definition of sympatric speciation and whether examples such as this are strictly sympatric in nature (Fitzpatrick et al., 2008), or instead are diverging in parapatry along clines. Nevertheless, in light of this it may be prudent to view allopatric and sympatric speciation not as opposing concepts, but rather as extremes of a spectrum on which all examples of divergence fall (Jiggins, 2006).

1.4 – *Chiloglanis*

The Family Mochokidae comprises nine genera and around 200 species of catfishes, and representatives can be found across the majority of sub-Saharan Africa (Vigliotta, 2008; Friel & Vigliotta, 2011). *Chiloglanis* Peters 1868 is the second most species rich genus of mochokids, second only to *Synodontis* Cuvier 1816 (Koblmüller et al., 2006; Seegers, 2008; Friel & Vigliotta, 2011; Day et al., 2013; Schmidt et al., 2014; 2015; 2016). Interestingly, a phylogenetic study into Mochokidae by Vigliotta (2008) found *Chiloglanis* is not monophyletic, but actually paraphyletic with respect to *Atopochilus*, *Atopodontus* and *Euchilichthys* (Friel & Vigliotta, 2011). To date 51 species of *Chiloglanis* have been described from freshwater systems across tropical Africa. One of the most characteristic features of the group is the presence of an oral disc (or sucker) formed from modifications of the jaws and lips, allowing them to support themselves while feeding within the fast-flowing river habitats that they occupy (Vigliotta, 2008; Friel & Vigliotta, 2011; Schmidt et al., 2014; 2015; 2016; 2017;). Despite the ease with which *Chiloglanis* individuals can be differentiated from other mochokid genera based on genus-typical traits such as pigmentation and mouth shape, *Chiloglanis* species are challenging to identify and differentiate (Seegers, 2008).

Chiloglanis species are relatively small in size with the largest individuals found at a standard length of ~10cm (Friel and Vigliotta, 2011; Schmidt et al., 2014; 2016; 2017). Individuals range in colour pattern but are often dark brown dorsally with interspersed cream patches (Friel and Vigliotta, 2011; Schmidt et al., 2015; 2017). Additionally, they are often speckled with dark brown melanophores (Friel and Vigliotta, 2011; Schmidt et al., 2015; 2017). This dark

dorsal colouration is likely a background matching strategy to aid in camouflaging individuals from the birds and larger fish that prey upon them.

Although found across South, East and West Africa, many species of *Chiloglanis* are highly endemic and are only found within single river systems (Seegers, 2008; Friel & Vigliotta 2011; Schmidt et al 2014; 2015; 2016; 2017). A few species, however, have much broader ranges. *Chiloglanis deckenii* Peters 1868 is found widely across Tanzania and Kenya, *Chiloglanis neumanni* Boulenger 1911 across Central and South Africa and *Chiloglanis batesii* Boulenger 1904 across Cameroon and the Congo Basin (Seegers, 2008). Due to the narrow distribution of the majority of described *Chiloglanis* species it is possible that samples from previous studies have been incorrectly identified as *C. deckenii*, *C. neumanni* or *C. batesii* as a result of near identical morphology (Seegers, 2008). In addition to the described species, there are a number of *Chiloglanis* species that are yet to be described (Seegers, 2008; Schmidt et al., 2014; 2015; 2016). Examples include *Chiloglanis* sp. "Ruvuma", found in the Muhowesi River in the Ruvuma drainage basin in Tanzania, and *Chiloglanis* sp. "Lupa", found in the Lupa River in the Lake Rukwa drainage basin in Tanzania (Seegers, 2008).

Numerous studies in the past have aimed to examine the geographic and genetic distribution of *Chiloganis* species throughout African freshwater systems and demonstrate their strong population structures and habitat specialisation. Morris et al. (2016) examined genetic structure of *Chiloglanis anoterus* Crass 1960, which occurs across South Africa, using cytochrome c oxidase 1 (CO1) and the control region (CR). Their findings, supported by nuclear amplified fragment length polymorphism (AFLP) data, suggest the existence of six separate clades within *C. anoterus*, with populations limited to much narrower ranges than the original widespread species, confirming the strong phylogeographic structure of *Chiloglanis* populations (Morris et al., 2016). Another analysis into mitochondrial DNA CO1 of *C. neumanni* collected from Eastern Zimbabwe Highlands freshwater ecoregion showed the existence of at least six groups of haplotypes within the region, with each group limited to narrow geographical ranges (Chakona et al., 2018). A study into *Chiloglanis brevibarbis* in Kenyan freshwaters using mitochondrial cytochrome b and supported by data from introns of the Growth Hormone nuclear gene exposed further unrecognised intraspecific divergence in the genus (Schmidt et al., 2014). Collectively, the evidence is supportive of extensive crypsis within *Chiloglanis* species, as well as the high levels of endemism.

1.5 – Aims of this study

This study builds on previous work testing for cryptic geographically structured diversity in *Chiloglanis* catfishes. The focus of the study is a region of East Africa (Malawi, northern Zambia and Southern Tanzania) that has a relatively poorly understood riverine fish fauna, and to date there have been no published studies that have attempted to clarify the taxonomic diversity of *Chiloglanis* present. The study also is the first to apply high-throughput sequencing to study phylogeographic structure in the genus. Specially this study aims to:

1) Generate a phylogenetic hypothesis for the populations of *Chiloglanis* catfishes across the study region.
2) Use the derived phylogeny to investigate whether multiple species are likely to be present in the study region.
3) Use the phylogeny to determine if catchment boundaries represent barriers to dispersal in the genus, and whether this isolation may have contributed to speciation in the genus.

Chapter Two

Uncovering phylogenetic diversity in riverine catfishes of East-Central Africa

Abstract

Despite the extensive biodiversity observed in freshwater species, increasing levels of anthropogenic activity and interference has led to freshwater fish species becoming some of the most threatened species on Earth. Large areas of global freshwater systems, however, remain largely unexplored and, as a result, species that inhabit these environments are often undescribed. Recent studies have not only been able to aid in the description of some of these species but have uncovered the presence of unrecognised divergence within previously described species. Due to the rigid spatial structure of many freshwater fish populations, it has been suggested that reproductive isolation as a result of geographic barriers to dispersal, and subsequently gene flow, is one of the key mechanisms behind speciation in these environments. To test whether catchments can act as geographic barriers to dispersal, and to investigate potential unrecognised biodiversity within freshwater fish populations, this study utilised double digestion restriction-site associated DNA (ddRAD) sequencing to provide data to map the phylogenetic relationship of *Chiloglanis* catfish samples collected from various catchments across Central and East Africa. The results confirm the strong endemic spatial distribution of *Chiloglanis* species consistent with their typical restriction to single catchments, in addition to identifying multiple clades in a region of which there is limited previous knowledge. These results confirm the hypothesis that there is unrecognised endemic diversity within *Chiloglanis* populations and the strict limitations to species ranges suggest that catchment boundaries may act as a geographic barrier preventing gene flow between neighbouring populations and driving the process of speciation.

2.1 – Introduction

Freshwater teleost fishes are believed to be among the most threatened of vertebrate taxa (Kadye & Moyo, 2008; Guo et al, 2019;) as a consequence of increasing levels of anthropogenic activity (Kadye & Moyo, 2008; Cooke et al., 2013; Guo et al, 2019). Among the factors identified as threats to freshwater fish populations are habitat fragmentation caused by the introduction of artificial geographic barriers (i.e. weirs and dams), the introduction of invasive species, and increasing volumes of pollution from both domestic and industrial sources (Rahel, 2007; Kadye & Moyo, 2008; Rashleigh et al. 2009; Cooke et al., 2013; Dias et al., 2013; Fuller et al., 2015). Despite making up only 0.01% of aquatic environments across the Earth by area, freshwater habitats contain 40% of all described fish species (Lundberg et al., 2000; Lévêque et al., 2008; Magurran et al., 2011; Bloom et al., 2013; Guo et al., 2019;). These fishes are of scientific importance and are used extensively to inform us of ecological and evolutionary processes that generate and maintain biodiversity. However, they have broader importance as a major source of protein for human diets worldwide, and as indicators of the intrinsic health of freshwater systems that are critical for providing drinking water and sustaining agriculture (Garcia, 2003; Cooke et al., 2013;). Therefore, conservation and management efforts are required to maintain freshwater fish species richness, density and distributions.

Knowledge of river fish ecology in Africa is extremely limited (Kadye & Moyo, 2008). This is illustrated by the rapid rate at which new species have been described in recent years, for example in catfish genus *Chiloglanis* (Friel & Vigliotta, 2011; Schmidt et al, 2015; 2017; Schmidt & Barrientos, 2019). *Chiloglanis* is the second largest genus within the family Mochokidae, a group of catfish found across the majority of sub-Saharan Africa, and contains 51 described species (Koblmüller et al., 2006; Seegers, 2008; Vigliotta; 2008; Friel & Vigliotta, 2011; Day et al., 2013; Schmidt et al., 2014; 2015; 2016). *Chiloglanis* species are highly habitat-specific, occupying fast-flowing rocky or reedy sections of streams and rivers. They are characterised by their suckermouth which helps in feeding on epilithic food resources, and also enables them to maintain position within their frequently turbulent environments (Engelbrecht & Mulder, 2000; Vigliotta, 2008; Friel & Vigliotta, 2011; Schmidt et al., 2014; 2015; 2016; 2017). Despite the genus being found widespread across sub-Saharan Africa, the majority of *Chiloglanis* species have very narrow distributions, often limited to a single catchment (Seegers, 2008; Friel & Vigliotta 2011; Schmidt et al 2014; 2015; 2016; 2017). In

addition to the 51 described species there are also numerous undescribed species, for example the Tanzanian taxa *C.* sp. *"Ruvuma"* and *C.* sp. *"Lupa"* found in the Ruvuma and Rukwa drainage basins respectively (Seegers, 2008).

Although the majority of *Chiloglanis* species have habitats limited to single river systems, there are a number of species which have much broader ranges. For example, *C. deckenii* is found across large areas of Tanzania and Kenya and *C. neumanni* is found across Central and South Africa (Seegers, 2008). However, due to the almost identical morphology of many *Chiloglanis* species and their typically narrow habitat restrictions, it is conceivable that these widely distributed species are multiple different "cryptic" species that have currently classified under the same name (Seegers, 2008). The plausibility of unrecognised diversity within widely distributed *Chiloglanis* species has been supported by several studies showing multiple genetically distinct and geographically separate populations within what has previously considered to be a single species (Schmidt et al., 2014; Morris et al., 2016; Chakona, 2018).

As genetic and phylogeographic study of fish develops, it is likely that an increasing diversity of freshwater fish will be recognised (Jousson et al., 2000; Bostock et al., 2006; Bickford et al., 2007; Lin et al., 2009). This study aimed to use next generation sequencing techniques to further investigate the unrecognised diversity among *Chiloglanis* populations, using samples collected from multiple Central and East African catchments. Data from double digest restriction-site associated DNA (ddRAD) sequencing was used to build a phylogenetic tree using a maximum likelihood approach, and clades were inferred using branch support from bootstrap analyses. The membership of clades was then considered from a biogeographic perspective. Ultimately, this study aims to provide key information to inform conservation and management of freshwater fish species and help to maintain species richness of East African *Chiloglanis* species.

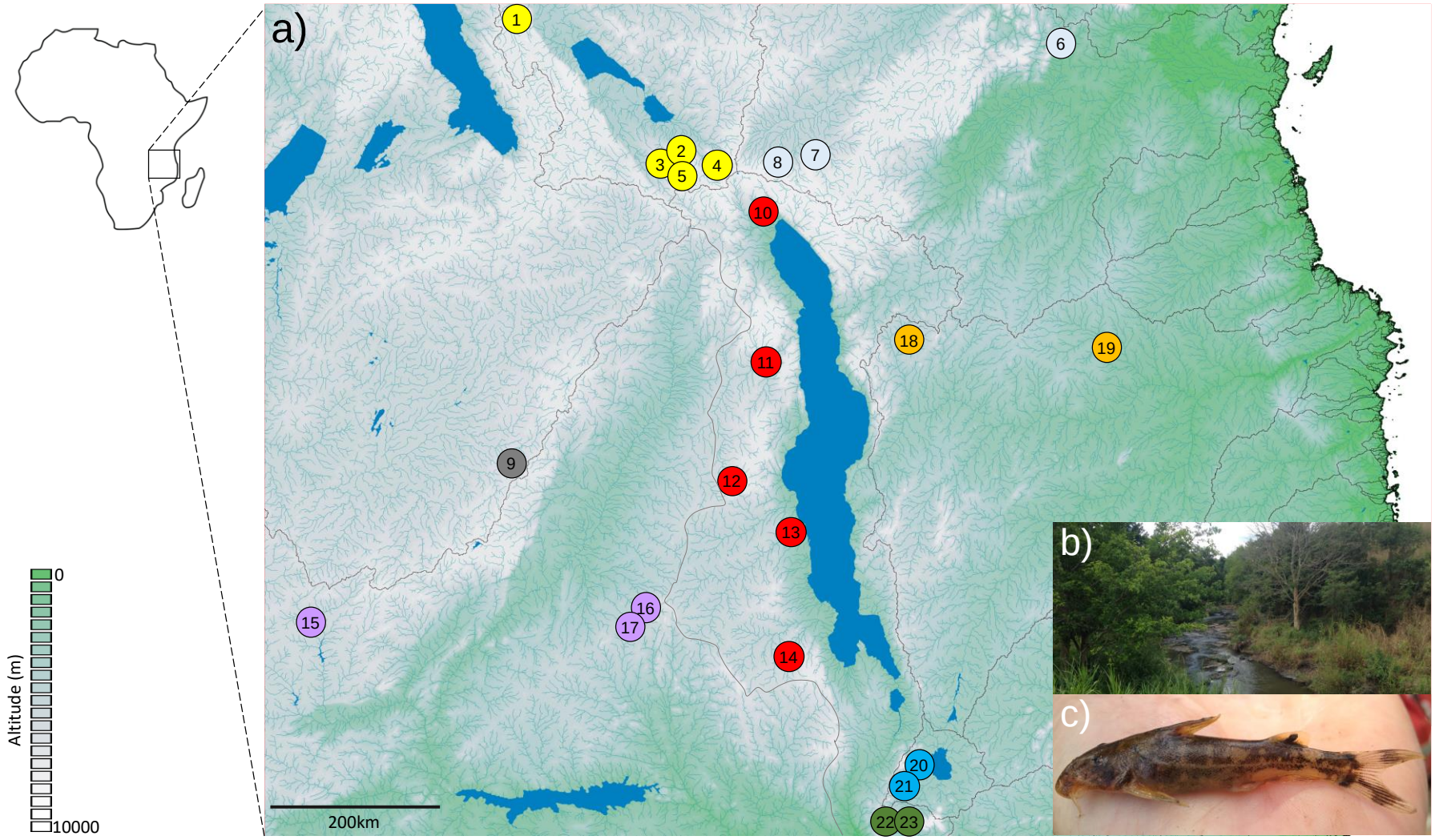

Figure 1. a) Collection sites for *Chiloglanis* specimens considered in this study. Colours indicate the sampling catchment. b) An example of typical *Chiloglanis* habitat (at Site 6 in the Rufiji/Ruaha system). C) *Chiloglanis* sp. (from Site 6 in the Rufiji/Ruaha system).

2.2 – Materials and Methods

Sample collection

Fish were collected from rocky or reedy sections of flowing streams and rivers using steel-framed D-shape handnets with 4mm black fryma mesh. These were placed downstream of a suitable substrate containing *Chiloglanis* specimens, and the substrate was then disturbed enabling the fish to enter the net. After collection fish were euthanised by an overdose of anaesthetic (clove oil) and preserved in absolute ethanol. Coordinates and of the sampling locations and their catchments are provided in Table 1. The sampling locations and associated catchments are shown Fig. 1. Collection permissions were provided the Department of Fisheries of the Government of Zambia, the Fisheries Research Unit of the Government of Malawi, and the Tanzania Commission for Science and Technology (COSTECH). Information regarding sample collection date and collectors can be found in Appendix 1.

Table 1. Geographic coordinates and river catchments of sampling events. and genetic clades recovered (Figure 1).

Site	Latitude °	Longitude °	Catchment	Clades recovered
1	-7.3097	31.0602	Rukwa	IX
2	-8.9141	32.8204	Rukwa	IX
3	-8.9132	32.8472	Rukwa	IX
4	-8.8994	33.3265	Rukwa	IX
5	-9.0348	32.9490	Rukwa	VIII, IX
6	-7.4848	37.0285	Rufiji	VI, VII
7	-8.7694	34.3748	Ruaha	V, VI
8	-8.8547	34.0861	Ruaha	V, XI
9	-12.1599	31.2318	Chambeshi	VIII, XI
10	-9.3954	33.8273	Malawi	IV
11	-11.0196	33.7857	Malawi	IV
12	-12.2722	33.4878	Malawi	IV
13	-12.8337	34.1623	Malawi	I
14	-14.1795	34.1245	Malawi	I
15	-13.7868	28.9997	Luangwa	III, XI
16	-13.7064	32.4897	Luangwa	VIII, XI
17	-13.7587	32.4498	Luangwa	III
18	-10.7017	35.3960	Rovuma	VII
19	-10.8473	37.4736	Rovuma	VII
20	-15.2791	35.4011	Chilwa	VII
21	-15.4860	35.2364	Chilwa	VII
22	-15.8463	35.1932	Ruo	II
23	-16.0004	35.3207	Ruo	II, X

DNA extraction and quantification

Genomic DNA extraction was undertaken using a modified version of the Wizard® genomic DNA extraction kit protocol (Promega, Madison USA). Small amounts of fin tissue were added to a 1.5 ml centrifuge tube containing 200µl of nuclei Lysis solution and 5µl of proteinase K (PK). Tubes were vortexed at high speed for 10 seconds before being incubated at 60°C for 30 minutes. 65µl of protein precipitation solution was then added to the tubes and they were vortexed at high speed before being centrifuged for 4 minutes at 14,600rpm. This produced a protein pellet overlain by supernatant containing the DNA. This supernatant was then poured into a new tube, and 200µl of cold 100% ethanol was added to precipitate the DNA. Tubes were then gently inverted 20 times and the tubes were centrifuged for 4 minutes at 14,600rpm to produce a pellet containing DNA. The supernatant was removed by pouring, and 200µl of cold 70% ethanol was added. Following another 4 min 14,600rpm centrifuge, the liquid was poured off and the tubes were left to air dry inverted on a paper towel. Finally, after the tubes had dried 30µl of H_2O was added to elute the DNA and the tubes were vortexed. DNA samples were then stored at 4°C, before purification with the QIAquick PCR Purification Kit (Qiagen, Hilden Germany). DNA concentration was measured using a dsDNA HS Assay Kit in a Qubit 4 fluorometer (ThermoFisher Scientific, Waltham USA) and samples were standardised to a concentration of 200ng of genomic DNA.

ddRAD library preparation and sequencing

Library preparation and sequencing was completed following the original double digest restriction enzyme associated (ddRAD) sequencing protocol (Peterson et al. 2012). Adapter stocks were annealed together and diluted to 0.4µM to produce barcoded adapters. A restriction-ligation was then used to ligate the barcoded adapters to the extracted DNA [15µl DNA, 5µl 1xCutSmart® Buffer (NEB, Ipswich USA), 0.5µl T4 ligase (NEB, Ipswich USA), 0.5µl ATP (NEB, Ipswich USA), 0.1µl EcoRI (NEB, Ipswich USA), 0.1µl MspI (NEB, Ipswich USA), 2µl adapter 1 (Appendix 2), 2µl adapter 2 (Appendix 2), 24.8µl H2O, PCR program: 3hr 37°C, 15min 68°C). The ligated product was amplified in four repeats of a 10-cycle PCR [4µl ligated DNA, 10µl 2xPhusion Flash PCR Master Mix (ThermoFisher Scientific, Waltham USA), 1µl primer 1 (Appendix 2), 1µl primer 2 (Appendix 2), 0.5µl BSA (NEB, Ipswich USA), 3.5µl H20, PCR program: 60s 98°C, x10(10s 98°C, 35s 55°C, 90s 72°C), 7 min 72°C). Amplified product was purified using the DNA Clean and ConcentratorTM Kit (Zymo Research, Irvine USA). Size selection was then completed using 1% agarose gel targeting DNA

in the range of 200-500bp in length and the QIAquick Gel Extraction Kit (Qiagen, Hilden Germany). Finally, libraries were quantified using the NEBNext Library Quant Kit (Illumina Inc., San Diego USA) to run a qPCR on a PCRmax Eco 48 (PCRmax, Stone UK). The library was processed at the Bristol Genomics Facility using the Illumina MiSeq v3 600cycle (2x300bp) kit yielding paired-end sequences each of 300bp read length. These sequences were first demultiplexed and adaptors trimmed using cutadapt v1.16 (Martin 2011), using the code in Appendix 3. Paired-end sequences were processed using ipyrad v0.7.25 (Eaton 2014), using code in Appendix 4.

Phylogenetic reconstruction

Phylogenies were reconstructed using both maximum likelihood (ML) and Bayesian inference (BI) approaches. A ML tree was constructed with phyML 3.0 (Guindon et al. 2010) using an online web server (http://www.atgc-montpellier.fr/phyml/). Automatic model selection by SMS (Lefort et al. 2017) was used (with the GTR model resolved as most likely), selecting the option to use the AIC (Akaike Information Criterion). The starting tree was as BIONJ tree, and branch support was estimated using the aLRT SH-like fast likelihood-based method. The BI tree was constructed using BEAST 2.6.2 (Bouckaert et al., 2014), using the GTR model. The run comprised 30 million generations, with one tree recorded every 1000 generations. The final tree and posterior probability branch support were calculated using TreeAnnotator within BEAST package, after the first 50% of trees were removed as burn-in. Both ML and BI trees were visualised using FigTree 1.4.4. (http://tree.bio.ed.ac.uk/software/figtree/) and rooted on the outgroup amphiliid catfish *Amphilius cf. uranoscopus.*

BI and ML methods were congruent in topologies, and the ML tree was used to display the results as ML approaches have been demonstrated to be reliable in other phylogenetic studies into fish populations using ddRAD sequencing methods (Beheregaray et al., 2017; Saenz-Agudelo et al., 2015; Umack et al., 2017). Following other ddRAD-based phylogenetic fish studies that use branch support values to identify previously unknown clades (Beheregaray et al., 2017; Umack et al., 2017), branches supported by both ML proportional support of > 0.7 and BI posterior probabilities of 1, were considered to represent putatively reproductively isolated taxa. The associations between the clade membership and sampling catchment were illustrated using an online Sankey Diagram Generator (http://sankey-diagram-generator.acquireprocure.com/).

The presence of a gap between the frequency distributions of interspecific (between clade) genetic variation and intraspecific (within clade) genetic variation can be used to help delimit taxa on a phylogeny. First, the ML phylogeny to estimate the "Patristic" genetic distance between individuals, using the cophenetic function in the ape package (Paradis and Schliep 2018) in R (R Core Team 2019). This matrix was then transformed to list using the dist2list function in spa (Zhang 2016). This list was then annotated to distinguish within-clade and between-clade genetic distances, which were plotted as a histogram using ggplot2 (Wickham 2016).

2.3 – Results

ddRAD sequencing

The final run resulted in 22,999,343 read pairs, of which 8,123,726 passed the filters and were assigned to individual samples. The final ipyrad output yielded 831 loci containing variable sites, and any samples yielding SNPs across less than 100 loci were removed from the dataset. The final analysed matrix resulted comprised 4119 variable positions and 65 individuals (64 *Chiloglanis* and 1 *A. cf. uranoscopus*). In the full dataset of 65 individuals, and 4119 positions, 105 were had an ambiguity code matching the consensus sequence. A total 3828 of these sites were parsimony informative, and there were 183 singletons. After removal of the outgroup *A. cf. uranoscopus*, 121 of the 4119 positions were either truly conserved or ambiguity code matching the consensus sequence. A total 3824 of these sites were parsimony informative, and there were 171 singletons.

Phylogenetic analysis

The ML phylogenetic tree, based on all 4119 SNPs and rooted on the outgroup *Amphilius cf. uranoscopus*, suggests the presence of 11 distinct *Chiloglanis* clades from across the eight catchments (Figure 2). Comparison of genetic distances both within and between described clades shows no overlap between interspecific and intraspecific genetic distance. This provides more support to our definition of 11 distinct clades (Figure 3).

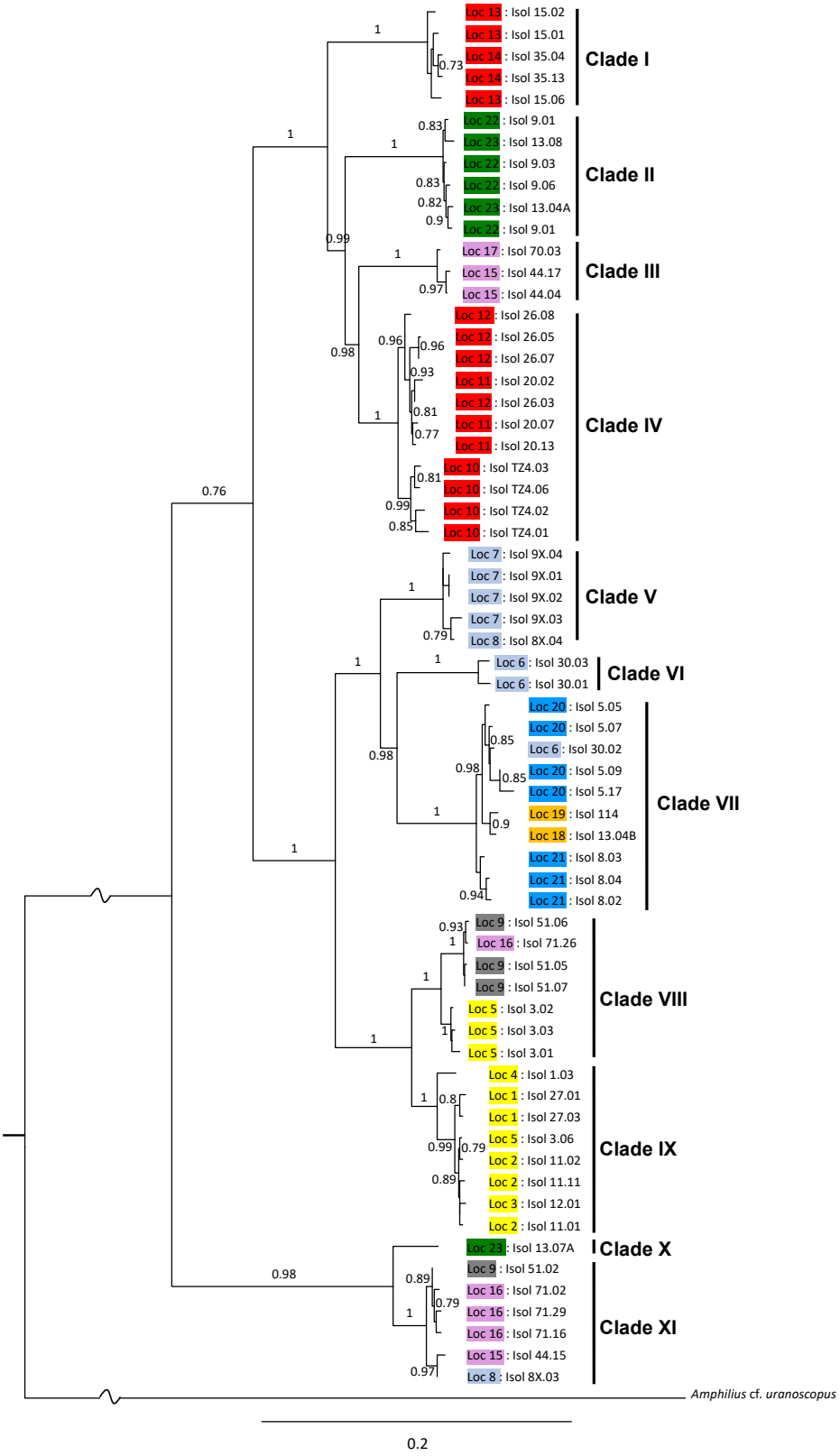

Figure 2. Maximum likelihood phylogeny based on 4119 SNPs. Numbers indicate proportional branch support values using the aLRT-SH-like approach in phyML. Scale bar represents the genetic distance on the basis of the GTR model. Asterisks represent branches with a Bayesian posterior probability (BPP) of 1 indicating maximum branch support.

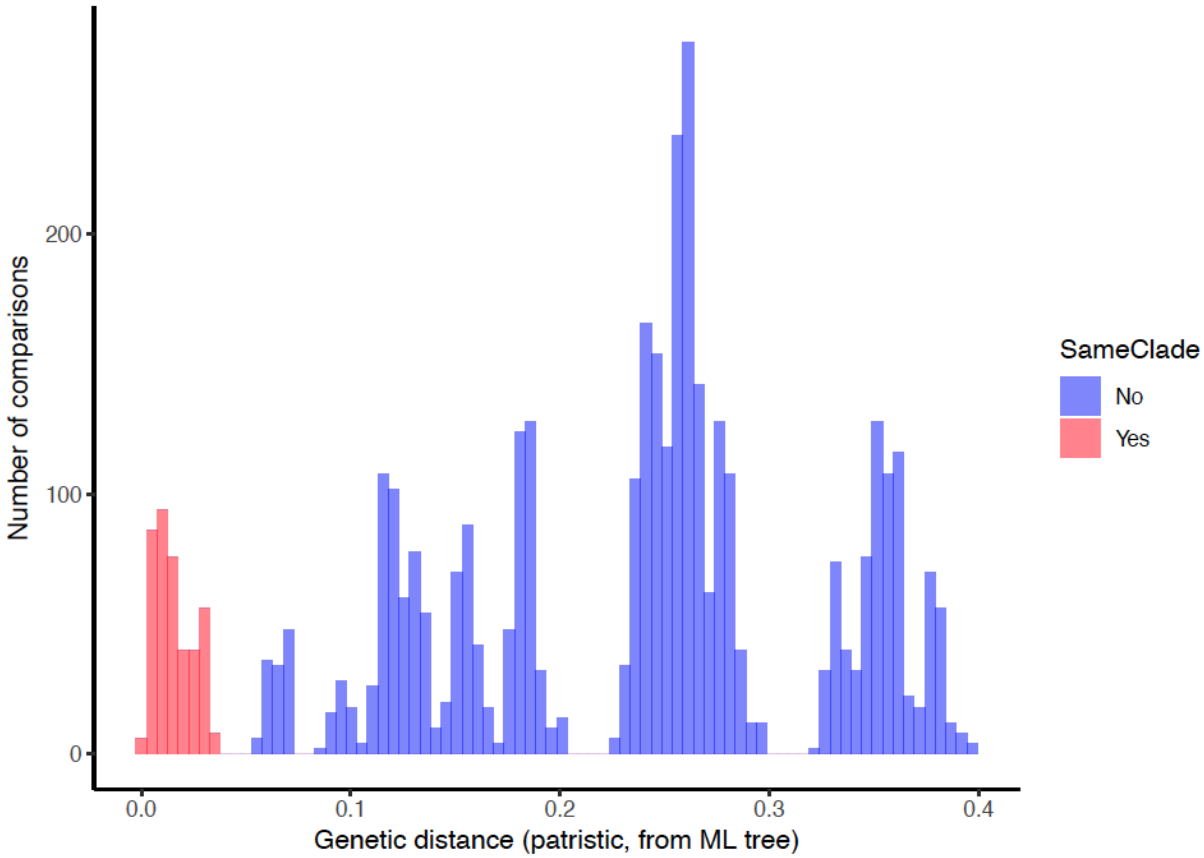

Figure 3. Frequency of genetic distances (patristic distances calculated from the ML tree; Figure 2) within clades (putative intraspecific variation) and between clades (putative interspecific variation).

Eight of the 11 clades were only recovered from single catchments and appear to be endemic. Of the three non-endemic clades, two contain samples from neighbouring catchments. Specifically, Clade VIII contains samples from the Chambeshi, Luangwa and Rukwa catchments, while Clade VII contains samples from the Ruaha, Rovuma and Chilwa catchments. Only one clade, Clade XI, had a disjunct distribution, being present in the from the Ruaha catchment as well as the Chambeshi and Luangwa catchments. Six of eight catchments from which samples were collected contain multiple genetically-distinct clades (Figure 4). We found some cases of individuals from different clades being recovered in sympatry (8 of 23 locations sampled), however more typically if multiple species were present in a catchment, they tended to be recovered in allopatry (Table 1).

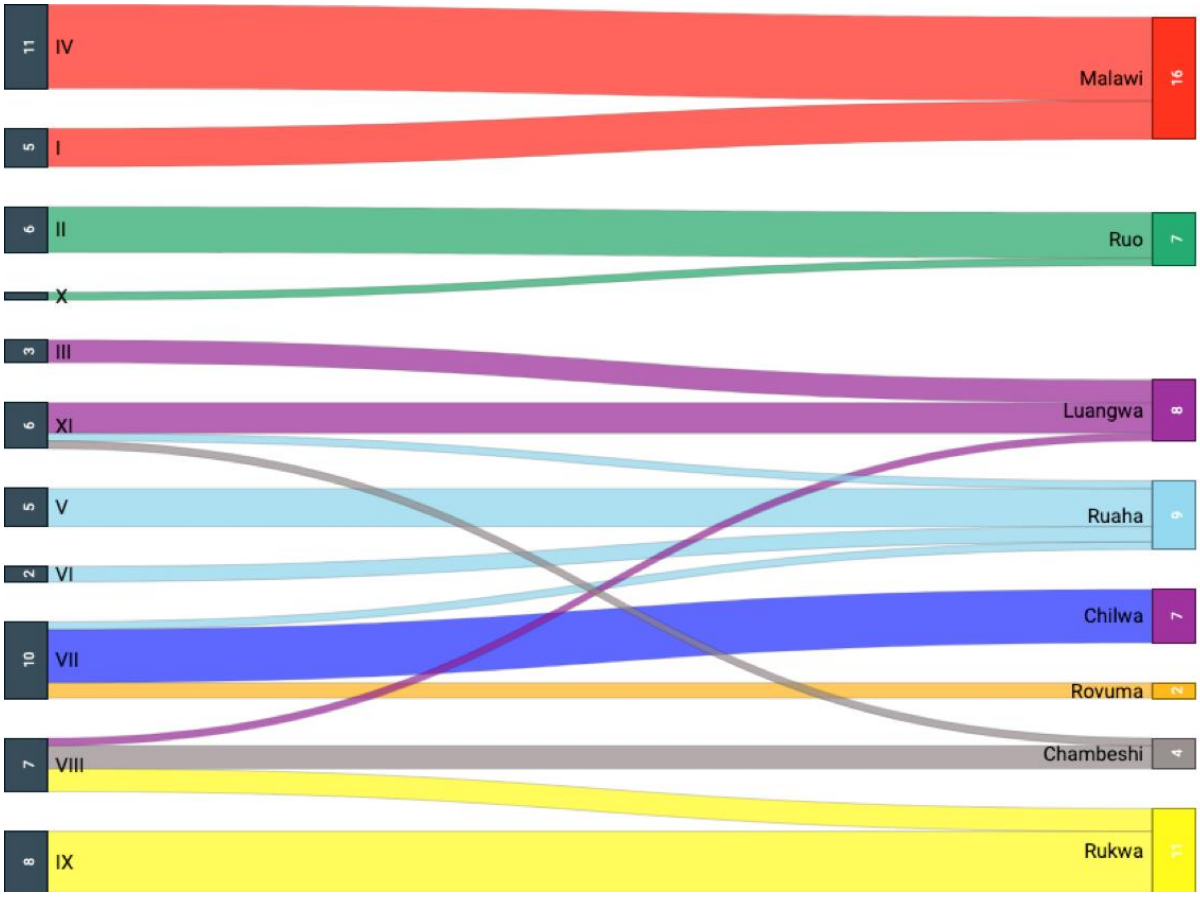

Figure 4. Sankey plot illustrating the catchments in which the 11 clades of *Chiloglanis* recovered in this study were found.

2.4 – Discussion

The results of phylogenetic analysis clearly demonstrate a high species richness within Central and East African *Chiloglanis* species. This is indicative of the unrecognised diversity as demonstrated for other regions in previous studies (Schmidt et al., 2014; Morris et al., 2016; Chakona, 2018). It is plausible that these represent reproductively isolated species, given the presence of multiple sympatric lineages across the phylogeny. While it is possible the 11 clades may indeed represent 11 distinct species, this may underestimate the species richness given clades could be further subdivided into allopatric lineages. For example, Clade IV could potentially be subdivided into two further allopatric sub-clades. Therefore, it is entirely possible that there may be even more unrecognised crypsis than can be confidently assessed by the data in this study.

On the other hand, the interpretation of 11 clades as 11 distinct species may overestimate diversity where allopatric sister clades may represent a single species. For example, Clade III is present in the Luangwa/Chambeshi catchments, while Clade IV is present in the Malawi catchment immediately adjacent the Luangwa. It is clear from this that further investigation into samples is needed in order to identify morphological traits that can help to determine the accuracy of lineage identification following the phylogenetic analysis in this study. Despite the difficulty with which *Chiloglanis* species are differentiated from each other based on morphological characteristics, jaw and tooth variation in particular have been identified as traits that can separate species (Seegers, 2008; Schmidt et al., 2014; Morris et al., 2016). Therefore, analysis of these morphological characteristics may help to provide clarity into the accuracy of clade identification. However, such traits are prone to plasticity and convergent evolution in lineages of fish under strong natural selection.

Relationships of sampled taxa to currently recognised species in the region

Here I considered only the phylogenetic relationships of the populations to gain a broad overview of the potential diversity of the region to inform further systematic investigation and taxonomic study. Such further study should accommodate presently described species and may require further field collections for genetic and morphological samples. Currently, six described species are present within the surrounding Lake Malawi and Lake Rukwa. Four species are known from the Lake Rukwa catchment, namely *Chiloglanis kalambo* Seegers 1996, *Chiloglanis mbozi* Seegers 1996, *Chiloglanis rukwaensis* Seegers 1996 and *Chiloglanis trilobatus* Seegers 1996. In our samples we recovered Clades VIII and IX from the Rukwa catchment, so therefore it plausible these specimens could be representative of one or more of these species. One further species reported from our study region is *Chiloglanis neumanii* Boulenger 1911 that has historically been reported to a broad distribution extending from the type locality in central Tanzania (Bubu River, near Dodoma), throughout southern and central Africa including Lake Malawi (Froese & Pauly 2019). Our samples cover a broad section of the reported range of this species, but we did not find evidence of the presence of a ubiquitous widely distributed species that incorporates population inside the Lake Malawi catchment. Therefore, our results support the conclusion of a recent genetic barcoding study of Southern African populations ascribed to this species that multiple cryptic species may be present (Chakona et al. 2018). Our results are also consistent with the assertion that the assignment of Lake Malawi catchment populations to *C. neumanni* should be revisited, as suggested on the

basis of morphological evidence (Konings & Tweddle 2019). A final described species that requires consideration in a taxonomic revision will be *Chiloglanis deckenii*, described from the Pangani river in northern Tanzania, but suggested to be present in the Ruaha and Rufiji systems. Therefore, *C. deckenii* could plausible be represented in our phylogeny by individuals from Clades V, VI or VII. In addition to the described species, several undescribed species are now mentioned from the region, including *C.* sp. "Ruvuma" (Seegers 1996) from the Ruvuma system (potentially our Clade VII), *C.* sp. "Lupa" (Seegers 1996) from the Rukwa system (potentially Clades VIII and IX) and *C.* sp. "Shire" (Chakona et al. 2018) reported from Lower Shire and Ruo systems of southern Malawi (potentially Clades II or X).

The results in this study are congruent with the key findings of other genetic studies of *Chiloglanis*, that tend to have uncovered previously unknown diversity. Schmidt et al. (2014) created a phylogeny from *Chiloglanis* samples collected across 16 sites in southern Kenya and northern Tanzania using sequence data from cytochrome b and introns from the nuclear growth hormone (GH) gene. From these data, they were able to uncover two previously undiscovered *Chiloglanis* clades each unique to a single river system (Schmidt et al., 2014). Another study by Morris et al. (2016) used mitochondrial DNA from cytochrome c oxidase 1 (CO1) and the control region (CR) from 117 *Chiloglanis* samples across 54 sampling sites to infer a phylogeny. This was supported by AFLP data from 307 variable regions. They were able to delineate six distinct clades from across these sites, the majority of which occurred in strict allopatry. In fact, there was only one site in the study at which two of the newly described clades were discovered in sympatry (Morris et al., 2016). Thus, the results in this study do share some similarities with these previous studies. Not only do they demonstrate a high species richness within Central and East African *Chiloglanis* species, they also provide evidence for the strong allopatric diversity within *Chiloglanis*. However, this study also uncovered a larger proportion of sympatric diversity than has been observed in similar recent studies. This may be due to the use of more loci in this study in comparison to that of previous studies, the majority of which rely on the use of mitochondrial data to infer phylogeny (Schmidt et al., 2014; Morris et al., 2016; Chakona, 2018). The use of nuclear DNA and ddRAD sequencing may have provided us with a wider view of *Chiloglanis* diversity and uncovered a larger rate of sympatric coexistence than has been witnessed in any previous study.

Allopatric diversity

Chiloglanis are habitat specialists and this is likely to drive the strong phylogeographic structure observed among species and populations (Seegers, 2008; Friel & Vigliotta 2011; Schmidt et al 2014; 2015; 2016; 2017). The results in this study demonstrate allopatric genetic diversity both between species in separate catchments, and between species present within different river systems in the same catchment. The results of this study, and others investigating phylogeographic structure in the genus, suggest that *Chiloglanis* species rarely cross catchment boundaries. However, when they do their distribution is almost always limited to multiple neighbouring catchments (Dignall, 1996; Seegers, 2008). Morris et al. (2016) demonstrated that population differentiation between *Chiloglanis* populations can take place even over short evolutionary timescales and narrow geographic scales in *Chiloglanis anoterus*. Specifically, it appears that this species is restricted to shallow waters and waterfall habitats and, and although waterfalls do not to act as a barrier to dispersal for this species, deep waters prevent their dispersal downstream from their high-altitude habitats (Morris et al., 2016). This evidence has similarities with work by Friel & Vigliotta (2011) on *Chiloglanis* species from opposing sides of Lake Tanganyika collected in the Malagarasi and Congo River basins. Despite the fact that the Congo and Malagarasi Rivers used to be connected before flooding lead to the formation of Lake Tanganyika ~20 million years ago, their results showed that no species of *Chiloglanis* observed in the study was found to exist in both drainage basins suggesting that Lake Tanganyika has served as a barrier preventing dispersal of *Chiloglanis* species which are restricted to riverine habitats (Friel & Vigliotta, 2011). The data obtained from these studies are consistent with the results of phylogenetic analysis in this study displaying strong levels of spatial structure in *Chiloglanis* populations and the existence of high levels of allopatric genetic divergence driven by geographic isolation.

Sympatric diversity

Despite the largely catchment-endemic nature of *Chiloglanis* species, the results demonstrate highlighted multiple cases where populations from distinct genetic lineages were living in sympatry. The existence of sympatric species is likely a result of secondary contact following events such as river captures or flooding, with populations remaining reproductively isolated due to the occurrence pre-existing isolating mechanisms (Morris et al., 2016; Schmidt et al., 2017). Evidence of sympatric coexistence of multiple taxa is relatively commonplace in our

results, for example representatives of Clades VIII and XI are both present in the Chambeshi and Luangwa catchments. These are not sister lineages, making secondary contact rather than sympatric speciation a more likely scenario for their coexistence (Schmidt et al., 2017). These results are important because most *Chiloglanis* clades we recovered were geographically restricted to single river systems, and therefore it is not readily clear if these represent good biological species that would be reproductively isolated on secondary contact, or instead allopatric populations of the same species. However, the discovery of sympatric populations with equivalent levels of sequence divergence provides stronger justification for the definition of clades in this study as distinct taxa, as these are certainly reproductively isolated.

It is possible that these species are able to coexist in sympatry as a result of niche partitioning and behavioural isolation. For example, Schmidt et al. (2016) displayed variation in microhabitat niche occupation in co-existing *Chiloglanis* species in West African rivers. In the Upper Niger River basin *Chiloglanis* cf. *micropogon* preferred faster flowing water and a rocky substrate, whereas C. sp. "Senegal/Niger" preferred slower water flow and was typically found in the presence of woody debris (Schmidt et al., 2016). This divergence in the occupation of niches within close proximity could key in the maintenance of diversity between sympatric species in our study. Additionally, very little is known about breeding habits of *Chiloglanis* species. It is possible, therefore, that mating behaviour may generate isolation between species in sympatry. In order to understand this further, we require more insight into the breeding behaviour of these species and how barriers to gene flow between co-existing species could be established.

2.5 – Concluding remarks

Considering the extent to which the majority of freshwater fish species are threatened, it is vital to gain further understanding into the levels of unrecognised diversity within freshwater populations. This study showed how high levels of unrecognised endemic biodiversity can be unveiled by widespread sampling and comparative phylogenetic analysis. As a result, it is reasonable to conclude that there may be a great deal more cryptic diversity across freshwater systems than is currently recognised. A greater understanding is required of habitat preference within riverine fish populations to help determine the factors that generate niche occupation. This will also help to determine the vulnerability of freshwater fish species to environmental

changes, considering the exponential increase of human activities that place pressures on African freshwaters. Further studies must be undertaken using both genetic and morphological data to give further insight into the levels of cryptic diversity among freshwater species. Additionally, this must be carried out over continental scales if the extent of diversity is to be fully catalogued and appreciated, and potential impact of anthropogenic factors fully evaluated. Considering the levels of unrecognised biodiversity shown in this study, and equivalent work, it is likely that many more freshwater populations are at risk than were originally thought.

References

Avise, J. C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7:371–379.

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Barton, N. and Bengtsson, B. O. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357–376.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. and Drummond, A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537.

Bermingham, E. and Martin, A. P. 1998. Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. Molecular Ecology 7:499–517.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K. and Das, I. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22:148–155.

Bloom, D. D., Weir, J. T., Piller, K. R. and Lovejoy, N. R. 2013. Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (Atherinopsidae). Evolution 67:2040–2057.

Bostock, B. M., Adams, M., Laurenson, L. J. B. And Austin, C. M. 2006. The molecular systematics of Leiopotherapon unicolor (Günther, 1859): testing for cryptic speciation in Australia's most widespread freshwater fish. Biological Journal of the Linnean Society 87:537–552.

Brown, D. M., Brenneman, R. A., Koepfli, K. P., Pollinger, J. P., Milá, B., Georgiadis, N. J., Louis, E. E., Grether, G. F., Jacobs, D. K. and Wayne, R. K. 2007. Extensive population genetic structure in the giraffe. BMC Biology 5:57.

Carreras-Carbonell, J., Macpherson, E. and Pascual, M. 2005. Rapid radiation and cryptic speciation in mediterranean triplefin blennies (Pisces: Tripterygiidae) combining multiple genes. Molecular Phylogenetics and Evolution 37:751–761.

Chakona, A., Kadye, W. T., Bere, T., Mazungula, D. N. and Vreven, E. 2018. Evidence of hidden diversity and taxonomic conflicts in five stream fishes from the Eastern Zimbabwe Highlands freshwater ecoregion. ZooKeys 768:69–95.

Cook, O. F. 1906. Factors of species-formation. Science, 23:506-507.

Cooke, S. J., Lapointe, N. W. R., Martins, E. G., Thiem, J. D., Raby, G. D., Taylor, M. K., Beard Jr, T. D. and Cowx, I. G. 2013. Failure to engage the public in issues related to inland fishes and fisheries: strategies for building public and political will to promote meaningful conservation. Journal of Fish Biology 83:997-1018.

Coyne J. A. and Orr H. A. 2004. Speciation. Sinauer Associates, Sunderland, Massachusetts.

Darwin, C. 2004. On the origin of species, 1859. Routledge.

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R. and Moritz, T. 2013. Continental Diversification of an African Catfish Radiation (Mochokidae: Synodontis). Systematic Biology 62:351–365.

Dias, M. S., Cornu, J.-F., Oberdorff, T., Lasso, C. A. and Tedesco, P. A. 2013. Natural fragmentation in river networks as a driver of speciation for freshwater fishes. Ecography 36:683–689.

Dignall, J. G. 1996. PlanetCatfish.com: The aquarium catfish website (https://www.planetcatfish.com/). [Electronic version accessed 15/11/2019].

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York.

Eaton, D. A. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30:1844-1849.

Engelbrecht, G. D. and Mulder, P. F. S. 2000. Allozyme variation in *Chiloglanis paratus* and *C. pretoriae* (Pisces, Mochokidae) from the Limpopo River system, Southern Africa. Water SA-Pretoria 26:111-114.

Finnilä, S., Hassinen, I., Ala-Kokko, L. and Majamaa, K. 2000. Phylogenetic Network of the mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the

Complete Coding Region by Conformation-Sensitive Gel Electrophoresis. The American Journal of Human Genetics 66:1017-1026.

Friel J. P. & Vigliotta, T. R. 2011. Three new species of African suckermouth catfishes, genus *Chiloglanis* (Siluriformes: Mochokidae), from the lower Malagarasi and Luiche rivers of western Tanzania. Zootaxa 3063:1–21.

Garcia, S. M. 2003. The ecosystem approach to fisheries: issues, terminology, principles, institutional foundations, implementation and outlook (No. 443). Food & Agriculture Organisation, Rome.

Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59:307-321

Guo, X.-Z., Zhang, G.-R., Wei, K.-J., Ji, W., Yan, R.-J., Wei, Q.-W. and Gardner, J. P. A. 2019. Phylogeography of the threatened tetraploid fish, Schizothorax waltoni, in the Yarlung Tsangpo River on the southern Qinghai-Tibet Plateau: implications for conservation. Scientific Reports 9:2704.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. and Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences 101:14812–14817.

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Molecular Ecology 10:537–549.

Hyde, J. R., Kimbrell, C. A., Budrick, J. E., Lynn, E. A. and Vetter, R. D. 2008. Cryptic speciation in the vermilion rockfish (*Sebastes miniatus*) and the role of bathymetry in the speciation process. Molecular Ecology 17:1122–1136.

Janzen, D. H., Burns, J. M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert, P. D. N. and Grishin, N. V. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. Proceedings of the National Academy of Sciences 114:8313–8318.

Jousson, O., Bartoli, P. and Pawlowski, J. 2000. Cryptic speciation among intestinal parasites (Trematoda: Digenea) infecting sympatric host fishes (Sparidae). Journal of Evolutionary Biology 13:778–785.

Knowles, L. L. and Maddison, W. P. 2002. Statistical phylogeography. Molecular Ecology 11:2623–2635.

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A. and Salzburger, W. 2006. Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis). BMC Evolutionary Biology 6:49.

Konings, A. & Tweddle, D. 2019. *Chiloglanis neumanni*. The IUCN Red List of Threatened Species 2019: e.T181738A58327132. Available at https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T181738A58327132.en

Kozak, G. M., Rudolph, A. B., Colon, B. L. and Fuller, R. C. 2012. Postzygotic Isolation Evolves before Prezygotic Isolation between Fresh and Saltwater Populations of the Rainwater Killifish, Lucania parva. International Journal of Evolutionary Biology 2012:1–11.

Le Gac, M., Hood, M. E., Fournier, E. and Giraud, T. 2007. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evolution 61:15–26.

Lefort, V., Longueville, J. E. & Gascuel, O. 2017. SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution 34:2422-2424.

Lemey, P., Rambaut, A., Drummond, A. J. and Suchard, M. A. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology 5:e1000520.

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M. L. J. and Tedesco, P. A. 2008. Global diversity of fish (Pisces) in freshwater. Hydrobiologia 595:545–567.

Lin, H. C., Sánchez-Ortiz, C. and Hastings, P.A. 2009. Colour variation is incongruent with mitochondrial lineages: cryptic speciation and subsequent diversification in a Gulf of California reef fish (Teleostei: Blennioidei). Molecular Ecology 18:2476–2488.

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A. and Willis, J. H. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. Philosophical Transactions of the Royal Society B: Biological Sciences 363:3009–3021.

Lundberg, J. G., Kottelat, M., Smith, G. R., Stiassny, M. L. J. and Gill, A. C. 2000. So Many Fishes, So Little Time: An Overview of Recent Ichthyological Discovery in Continental Waters. Annals of the Missouri Botanical Garden 87:26.

Machordom, A. and Doadrio, I. 2001. Evidence of a Cenozoic Betic–Kabilian Connection Based on Freshwater Fish Phylogeography (Luciobarbus, Cyprinidae). Molecular Phylogenetics and Evolution 18:252–263.

Magurran, A. E., Khachonpisitsak, S. and Ahmad, A. B. 2011. Biological diversity of fish communities: pattern and process. Journal of Fish Biology 79:1393–1412.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17:10-12.

Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

Mayr E. 1963. Animal species and evolution. Harvard University Press, Cambridge, Massachusetts.

Morris, J., Ford, A. G. P., Ali, J. R., Peart, C. R., Bills, R. and Day, J. J. 2016. High levels of genetic structure and striking phenotypic variability in a sexually dimorphic suckermouth catfish from the African Highveld. Biological Journal of the Linnean Society 117:528–546.

Moyle, L. C. 2007. Comparative Genetics of Potential Prezygotic and Postzygotic Isolating Barriers in a Lycopersicon Species Cross. Journal of Heredity 98:123–135.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. and Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences 98:12084–12088.

Nosil, P., Vines, T. H. and Funk, D. J. 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719.

Palumbi, S. R. 1994. Genetic Divergence, Reproductive Isolation, and Marine Speciation. Annual Review of Ecology and Systematics 25:547–572.

Peterson, B., Weber, J., Kay, E., Fisher, H. and Hoekstra, H. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE, 7:e37135.

Rahel, F. J. 2007. Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. Freshwater Biology 52:696–710.

Rashleigh, B., Hardwick, D. and Roux, D. 2009. Fish assemblage patterns as a tool to aid conservation in the Olifants River catchment (East), South Africa. Water SA 35.

Roca, A.L., Georgiadis, N., Pecon-Slattery, J., and O'Brien, S. J. 2001. Genetic Evidence for Two Species of Elephant in Africa. Science 293:1473–1477.

Rosser, N., Freitas, A. V. L., Huertas, B., Joron, M., Lamas, G., Mérot, C., Simpson, F., Willmott, K. R., Mallet, J. and Dasmahapatra, K. K. 2018. Cryptic speciation associated with geographic and ecological divergence in two Amazonian Heliconius butterflies. Zoological Journal of the Linnean Society 186:233–249.

Schmidt, R. C. and Barrientos, C. 2019. A new species of suckermouth catfish (Mochokidae: *Chiloglanis*) from the Rio Mongo in Equatorial Guinea. Zootaxa 4652:507–519.

Schmidt, R. C., Bart Jr, H. L. and Nyingi, W. D. 2015. Two new species of African suckermouth catfishes, genus *Chiloglanis* (Siluriformes: Mochokidae), from Kenya with remarks on other taxa from the area. Zootaxa 4044:045-064.

Schmidt, R. C., Bart, H. L. and Pezold, F. 2016. High levels of endemism in suckermouth catfishes (Mochokidae: *Chiloglanis*) from the Upper Guinean forests of West Africa. Molecular Phylogenetics and Evolution 1001:199–205.

Schmidt, R. C., Bart, H. L., Nyingi, D. W. and Gichuki, N. N. 2014. Phylogeny of suckermouth catfishes (Mochokidae: *Chiloglanis*) from Kenya: The utility of Growth Hormone introns in species level phylogenies. Molecular Phylogenetics and Evolution 79:415–421.

Schmidt, R. C., Bart, H. L., Pezold, F. and Friel, J. P. 2017. A Biodiversity Hotspot Heats Up: Nine New Species of Suckermouth Catfishes (Mochokidae: *Chiloglanis*) from Upper Guinean Forest Streams in West Africa. Copeia 105:301–338.

Schwartz, A. K., Weese, D. J., Bentzen, P., Kinnison, M. T. and Hendry, A. P. 2010. Both Geography and Ecology Contribute to Mating Isolation in Guppies. PLoS ONE 5:e15659.

Searle, J. 2000. Phylogeography - The History and Formation of Species. Heredity 85:201.

Seegers L. 2008. The catfishes of Africa. A handbook for identification and maintenance. Aqualog Verlag A.C.S. GmbH., Rodgau.

Sivasundar, A., Bermingham, E. and Orti, G. 2001. Population structure and biogeography of migratory freshwater fishes (Prochilodus: Characiformes) in major South American rivers. Molecular Ecology 10:407–417.

Sutra, N., Kusumi, J., Montenegro, J., Kobayashi, H., Fujimoto, S., Masengi, K. W., Nagano, A. J., Toyoda, A., Matsunami, M., Kimura, R. and Yamahira, K. 2019. Evidence for sympatric speciation in a Wallacean ancient lake. Evolution 73:1898-1915.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. and Cosson, J. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7:453–464.

Theodorakis, C. W., Lee, K.-L., Adams, S. M. and Law, C. B. 2006. Evidence of Altered Gene Flow, Mutation Rate, and Genetic Diversity in Redbreast Sunfish from a Pulp-Mill-Contaminated River. Environmental Science & Technology 40:377–386.

Tiercelin, J. J. & Mondeguer, A. 1991. The geology of the Tanganyika Trough. In: Lake Tanganyika and its life (ed. G. W. Coulter) 7-48. Oxford University Press.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. and Wallace, D. C. 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835-1850.

Verheyen, E., Rüber, L., Snoeks, J. and Meyer, A. 1996. Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351:797-805.

von der Heyden, S., Bowie, R. C. K., Prochazka, K., Bloomer, P., Crane, N. L. and Bernardi, G. 2011. Phylogeographic patterns and cryptic speciation across oceanographic barriers in South African intertidal fishes. Journal of Evolutionary Biology 24:2505–2519.

Vigliotta, T. R. 2008. A phylogenetic study of the African catfish family Mochokidae (Osteichthyes, Ostariophysi, Siluriformes), with a key to genera. Proceedings of the Academy of Natural Sciences of Philadelphia 157:73–136.

Wallace, D. 1994. Mitochondrial DNA sequence variation in human evolution and disease. Proceedings of the National Academy of Sciences 91:8739-8746.

Appendix 1 – Sample collection information

Sample	Date	Site	Collectors
27.01	30/07/17	1	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
27.03	30/07/17	1	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
11.01	25/07/17	2	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
11.02	25/07/17	2	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
11.11	25/07/17	2	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
12.01	25/07/17	3	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
1.03	01/09/12	4	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
OG3	01/09/12	4	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
3.01	23/07/17	5	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
3.02	23/07/17	5	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
3.03	23/07/17	5	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
3.06	23/07/17	5	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
30.01	02/08/17	6	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
30.02	02/08/17	6	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
30.03	02/08/17	6	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge
9X.01	04/09/12	7	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
9X.02	04/09/12	7	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
9X.03	04/09/12	7	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
9X.04	04/09/12	7	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
8X.03	04/09/12	8	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
8X.04	04/09/12	8	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
51.02	17/06/10	9	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
51.05	17/06/10	9	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
51.06	17/06/10	9	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
51.07	17/06/10	9	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
TZ4.01	16/07/11	10	Martin Genner, Benjamin Ngatunga, George Turner
TZ4.02	16/07/11	10	Martin Genner, Benjamin Ngatunga, George Turner
TZ4.03	16/07/11	10	Martin Genner, Benjamin Ngatunga, George Turner
TZ4.06	16/07/11	10	Martin Genner, Benjamin Ngatunga, George Turner
20.02	26/05/10	11	Jennifer Swanstrom, Denis Tweddle
20.07	26/05/10	11	Jennifer Swanstrom, Denis Tweddle
20.13	26/05/10	11	Jennifer Swanstrom, Denis Tweddle
26.03	29/05/10	12	Jennifer Swanstrom, Denis Tweddle
26.05	29/05/10	12	Jennifer Swanstrom, Denis Tweddle
26.07	29/05/10	12	Jennifer Swanstrom, Denis Tweddle
26.08	29/05/10	12	Jennifer Swanstrom, Denis Tweddle
15.01	23/05/10	13	Jennifer Swanstrom, Denis Tweddle
15.02	23/05/10	13	Jennifer Swanstrom, Denis Tweddle
15.06	23/05/10	13	Jennifer Swanstrom, Denis Tweddle
35.04	09/06/10	14	Martin Genner, Jennifer Swanstrom
35.13	09/06/10	14	Martin Genner, Jennifer Swanstrom
44.04	15/06/10	15	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
44.15	15/06/10	15	Martin Genner, Cyprian Katongo, Jennifer Swanstrom
44.17	15/06/10	15	Martin Genner, Cyprian Katongo, Jennifer Swanstrom

Appendix 1 continued – Sample collection information

Sample	Date	Site	Collectors
71.02	12/07/10	16	Martin Genner, Jennifer Swanstrom
71.16	12/07/10	16	Martin Genner, Jennifer Swanstrom
71.26	12/07/10	16	Martin Genner, Jennifer Swanstrom
71.29	12/07/10	16	Martin Genner, Jennifer Swanstrom
70.03	20/06/10	17	Martin Genner, Jennifer Swanstrom
13.04 (B)	06/09/12	18	Martin Genner, Benjamin Ngatunga, Alan Smith, Jennifer Swanstrom, George Turner
114	17/08/13	19	Martin Genner, Benjamin Ngatunga, Asilatu Shechonge, Alan Smith
5.05	15/05/10	20	Jennifer Swanstrom
5.07	15/05/10	20	Jennifer Swanstrom
5.09	15/05/10	20	Jennifer Swanstrom
5.17	15/05/10	20	Jennifer Swanstrom
8.02	16/05/10	21	Jennifer Swanstrom
8.03	16/05/10	21	Jennifer Swanstrom
8.04	16/05/10	21	Jennifer Swanstrom
9.01	17/05/10	22	Jennifer Swanstrom
9.02	17/05/10	22	Jennifer Swanstrom
9.03	17/05/10	22	Jennifer Swanstrom
9.06	17/05/10	22	Jennifer Swanstrom
13.04 (A)	19/05/10	23	Jennifer Swanstrom
13.07 (A)	19/05/10	23	Jennifer Swanstrom
13.08 (A)	19/05/10	23	Jennifer Swanstrom

Appendix 2 – Adapter barcodes and PCR indices.

Adapter stock barcode	
GCATG_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATG-3'
AACCA_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCA-3'
CGATC_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATC -3'
TCGAT_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGAT -3'
TGCAT_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCAT -3'
CAACC_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACC -3'
GGTTG_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTG -3'
AAGGA_EcoRI_1.1	5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGA -3'
GCATG_EcoRI_1.2	5'-[PHO]AATTCATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -3'
AACCA_EcoRI_1.2	5'-[PHO]AATTTGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CGATC_EcoRI_1.2	5'-[PHO]AATTGATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TCGAT_EcoRI_1.2	5'-[PHO]AATTATCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TGCAT_EcoRI_1.2	5'-[PHO]AATTATGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CAACC_EcoRI_1.2	5'-[PHO]AATTGGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
GGTTG_EcoRI_1.2	5'-[PHO]AATTCAACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
AAGGA_EcoRI_1.2	5'-[PHO]AATTTCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
MspI_2.1	5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
MspI_2.2	5'-[PHO]CGAGATCGGAAGAGCGAGAACAA-3'
Primer Index	
PCR1	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3'
PCR2_Idx_1	5'-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_2	5'-CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_3	5'-CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_4	5'-CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_5	5'-CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_6	5'-CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_7	5'-CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_8	5'-CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_9	5'-CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_10	5'-CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_11	5'-CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC-3'
PCR2_Idx_12	5'-CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC-3'

Appendix 3 – Code for trimming and demultiplexing.

#Contents of file PCRIndex.txt

```
>PCR2_Idx_1_ATCACG
GTCACATCACGATCTC
>PCR2_Idx_2_CGATGT
GTCACCGATGTATCTC
>PCR2_Idx_3_TTAGGC
GTCACTTAGGCATCTC
>PCR2_Idx_4_TGACCA
GTCACTGACCAATCTC
>PCR2_Idx_5_ACAGTG
GTCACACAGTGATCTC
>PCR2_Idx_6_GCCAAT
GTCACGCCAATATCTC
>PCR2_Idx_7_CAGATC
GTCACCAGATCATCTC
>PCR2_Idx_8_ACTTGA
GTCACACTTGAATCTC
>PCR2_Idx_9_GATCAG
GTCACGATCAGATCTC
>PCR2_Idx_10_TAGCTT
GTCACTAGCTTATCTC
>PCR2_Idx_11_GGCTAC
GTCACGGCTACATCTC
>PCR2_Idx_12_CTTGTA
GTCACCTTGTAATCTC
```

```
#Contents of the file AdaptorIndex.txt


>Adaptor1_CATGC
AATTCATGCAGA
>Adaptor2_TGGTT
AATTTGGTTAGA
>Adaptor3_GATCG
AATTGATCGAGA
>Adaptor4_ATCGA
AATTATCGAAGA
>Adaptor5_ATGCA
AATTATGCAAGA
>Adaptor6_GGTTG
AATTGGTTGAGA
>Adaptor7_CAACC
AATTCAACCAGA
>Adaptor8_TCCTT
AATTTCCTTAGA
```

Appendix 3 continued – Code for trimming and demultiplexing.

```
#Cutadapt code for step 1 of demultiplex (Index matching), separating out
the raw .fastq paired end reads into 12 pairs of files each corresponding
to one PCR index. Code assumes you are in the folder, and PCRIndex.txt and
fastq raw data files with the appropriate names are present.

cutadapt --no-trim -a file:PCRIndex.txt -o trimmed-{name}.1 -p trimmed-
{name}.2 RC-ddRAD_S1_L001_R1_001.fastq RC-ddRAD_S1_L001_R2_001.fastq

#Trim the adaptors from the reads. NB edit the name of the input files each
time. Do this 12 times, one for each index.

cutadapt -a CGAGATCGGAAGAGC -o Index1_trim-{name}_R1_.fastq -p Index1_trim-
{name}_R2_.fastq trimmed-PCR2_Idx_1_ATCACG.1 trimmed-PCR2_Idx_1_ATCACG.2


#Each pair of trimmed files (one for each index) then placed into their own
folder, with a copy of AdaptorIndex.txt. NB edit the name of the input
files each time. Do this 12 times, once for each index.

cutadapt -a file:AdaptorIndex.txt -o Index1_R1_-{name}.fastq -p Index1_R2_-
{name}.fastq Index1_trim-1_R2_.fastq Index1_trim-1_R1_.fastq
```

Appendix 4 – Code for Ipyrad v0.7.25.

```
#Structure of the ipyrad file. Note all fastq files for analysis were placed in the folder ./Test/*.fastq

------- ipyrad params file (v.0.7.25)-------------------------------------------
ip_MG_Henry                            ## [0] [assembly_name]: Assembly name. Used to name output directories for assembly steps
./                                      ## [1] [project_dir]: Project dir (made in curdir if not present)
                                        ## [2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
                                        ## [3] [barcodes_path]: Location of barcodes file
./Test/*.fastq                          ## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
denovo                                 ## [5] [assembly_method]: Assembly method (denovo, reference, denovo+reference, denovo-reference)
                                       ## [6] [reference_sequence]: Location of reference sequence file
pairddrad                               ## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
                                        ## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1,cut2)
5                                       ## [9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
33                                      ## [10] [phred_Qscore_offset]: phred Q score offset (33 is default and very standard)
5                                       ## [11] [mindepth_statistical]: Min depth for statistical base calling
3                                       ## [12] [mindepth_majrule]: Min depth for majority-rule base calling
10000                                   ## [13] [maxdepth]: Max cluster depth within samples
0.85                                    ## [14] [clust_threshold]: Clustering threshold for de novo assembly
0                                       ## [15] [max_barcode_mismatch]: Max number of allowable mismatches in barcodes
0                                       ## [16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
10                                      ## [17] [filter_min_trim_len]: Min length of reads after adapter trim
2                                       ## [18] [max_alleles_consens]: Max alleles per site in consensus sequences
5, 5                                    ## [19] [max_Ns_consens]: Max N's (uncalled bases) in consensus (R1, R2)
8, 8                                    ## [20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus (R1, R2)
32                                      ## [21] [min_samples_locus]: Min # samples per locus for output
20, 20                                  ## [22] [max_SNPs_locus]: Max # SNPs per locus (R1, R2)
8, 8                                    ## [23] [max_Indels_locus]: Max # of indels per locus (R1, R2)
0.5                                     ## [24] [max_shared_Hs_locus]: Max # heterozygous sites per locus (R1, R2)
0, 0, 5, 0                              ## [25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see docs)
0, 0, 0, 0                              ## [26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2)
p, s,                                   ## [27] [output_formats]: Output formats (see docs)
                                        ## [28] [pop_assign_file]: Path to population assignment file
```

Appendix 5 – Bayesian phylogenetic tree of the full dataset, including posterior probability branch support